PSY 652: Lab Exercises for Module 2

Answer Key

Power Analysis: Two different methods

Method 1: Use df_{hyp} & df_{err}

Definitions:

df_{hyp}: Is basically an index of the complexity of the problem. For example, when comparing the 2 groups, $df_{hyp} = 1$. When comparing 5 groups, $df_{hyp} = 4$. If you are evaluating 7 predictors of an outcome, your $df_{hyp} = 6$.

df_{err}: The degrees of freedom for the estimate of error used in the test.

Sample size: $N = df_{hyp} + df_{err} + 1$

1. Use Murphy, Myors & Wolach (2014) Appendix E to compute sample size for the following problems at a power level of .80 and α = .05. This table uses two alternative ways of indexing effect size, *d* and *PV*, or the percentage of variance explained. In general, *PV* = .01, .10 and .25 represent small, moderately large and large effects.

    a. For a study comparing 2 groups, calculate the df_{err} & sample size required to detect a PV of .01

    df_{err} = 775, N = 777

    b. For a study examining 4 predictors of an outcome, calculate the df_{err} & sample size required to detect a PV of .10

    df_{err} = 101, N = 105

    c. For a study comparing 8 groups, calculate the df_{err} & sample size required to detect a PV of .25

    df_{err} = 52 , N = 58 (use *PV* = 0.24*)*

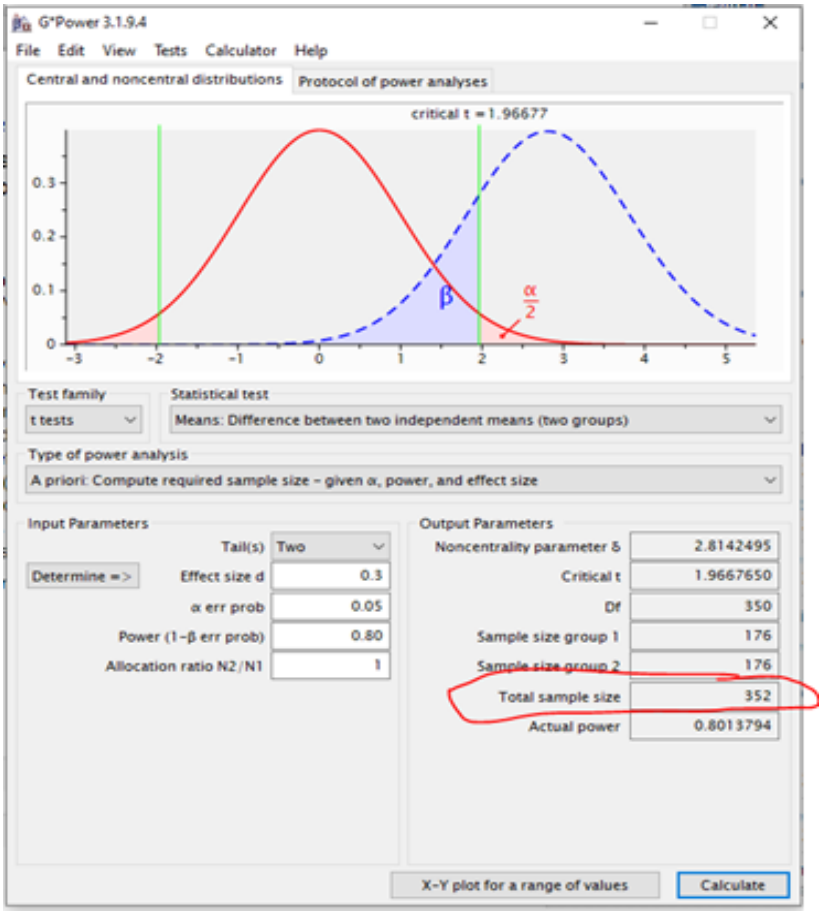
Method 2: Use G*Power to perform *a priori*, *compromise,* and *post-hoc* power analyses

G*Power can be freely downloaded at:
http://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html

Use G*power to answer the following questions:

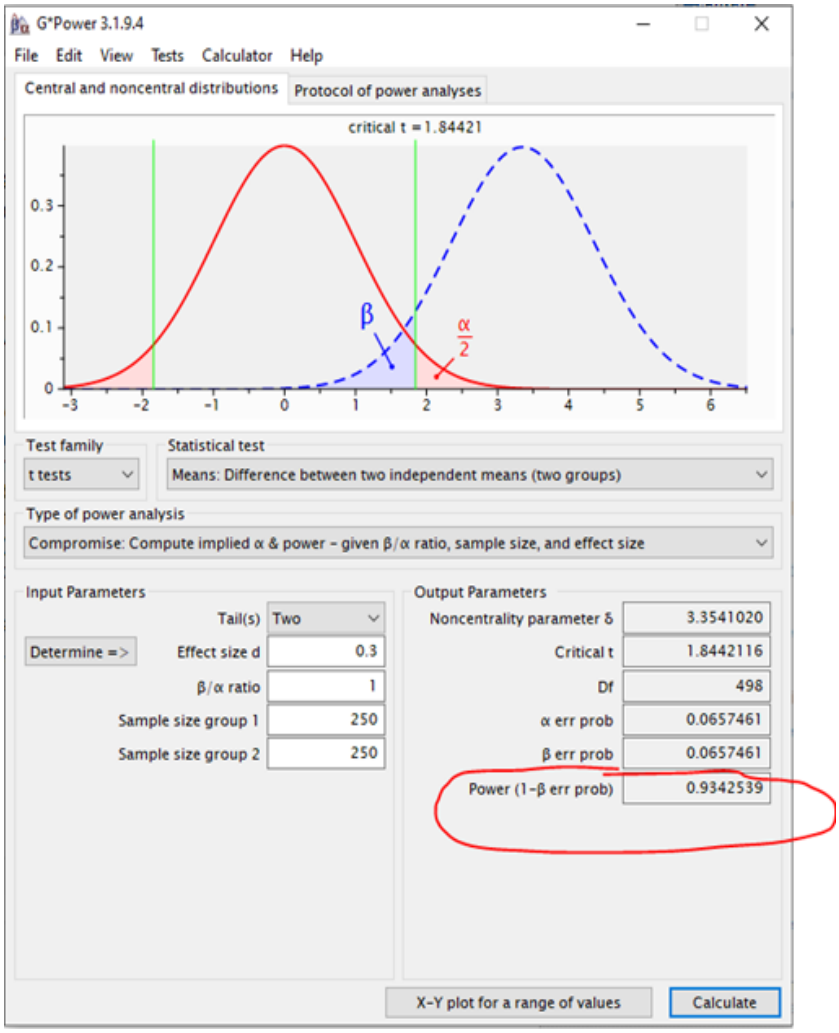
1) Your advisor has recently found themselves in a busy mess and they don't have time to run a power analysis for a grant that is due *tonight!* Because you are an ambitious grad student, you take on the task. The study is a randomized control treatment assessing the mean differences of anxiety scores between participants that took a 12-week mindfulness meditation class vs. those in a control group. An independent samples t-test will be used to assess this difference. You are instructed to use an alpha level of .05 and a recent meta-analysis that came out has shown that mindfulness interventions and anxiety tend to show an effect size of 0.30. What sample size would be needed to achieve a power level of .80?

Answer: A total of 352 participants are needed to get a power level of .80



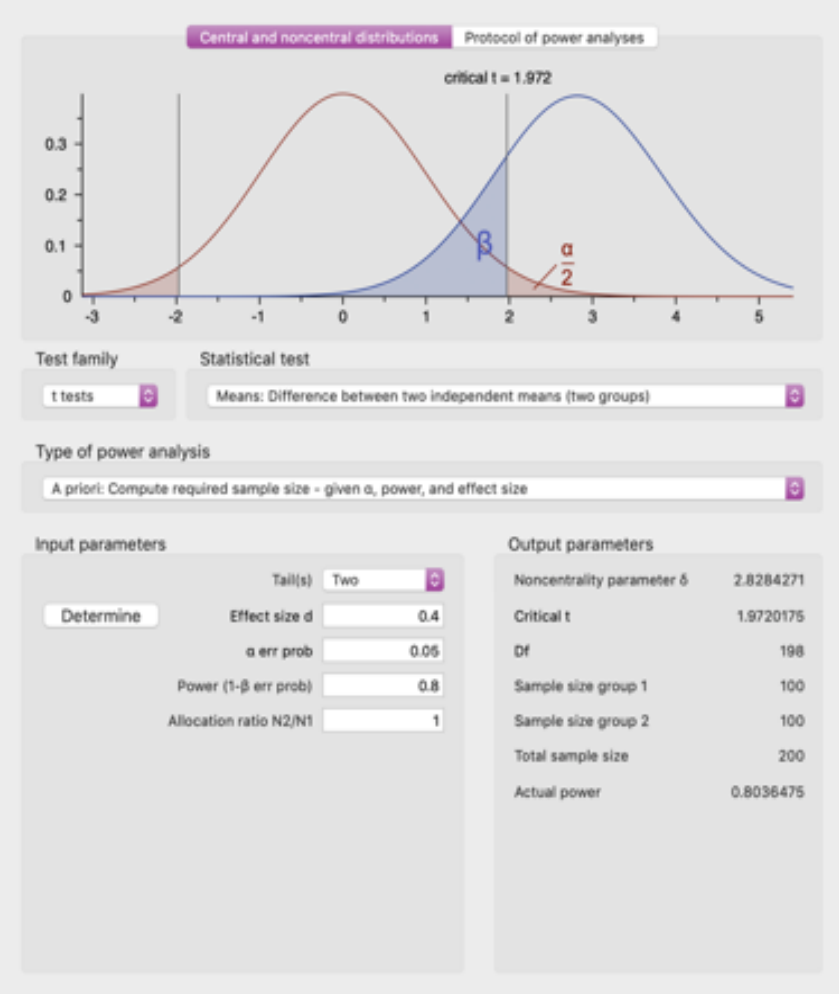
1a) After giving your advisor your splendid power analysis results, they tell you that they anticipate having 500 participants (250 in each condition). Using the same alpha level and effect sizes as before, what is the power of this study?

Answer: Power = .93



2) You and your fellow graduate students are planning to conduct a study to compare the average overall GPA of undergraduate students in two different majors at CSU: psychology and physics. Given a desired power level of .80 and an alpha threshold of .05, what sample size would you need to determine if an effect size of 0.40 exists, in *any* direction, between the mean GPA of students in each major? Assume that the number of students from each major are equal in your sample.

Answer: You would need an N = 200 to detect this effect.



2a) In a sentence or two, what does an effect size of d = 0.40 *mean*, in terms of the average effect of being in one major versus the other on overall GPA? (hint: consider the formula for Cohen's *d*)

> *"Cohen's d = (Mt - Mc)/SD, where Mt and Mc are the treatment and control group means, respectively, and SD is the pooled standard deviation" (Murphy, Myors, & Wolach, 2014).*

Possible answers:

- ***d* = 0.40 indicates that the GPA means of the two majors differ by 2/5 of a standard deviation.**
- **The difference in the average GPA of students in one major versus the other is 40% as large as the SD of GPA (the outcome measure) within each of the majors.**
- **The average effect of major is 2/5ths as large as the variability in GPA among students within the same major.**

2b) If the study has sufficient power to detect an effect of d = 0.40, what is the full range of effect sizes that can it reliably detect?

Answer: $d \geq 0.40$

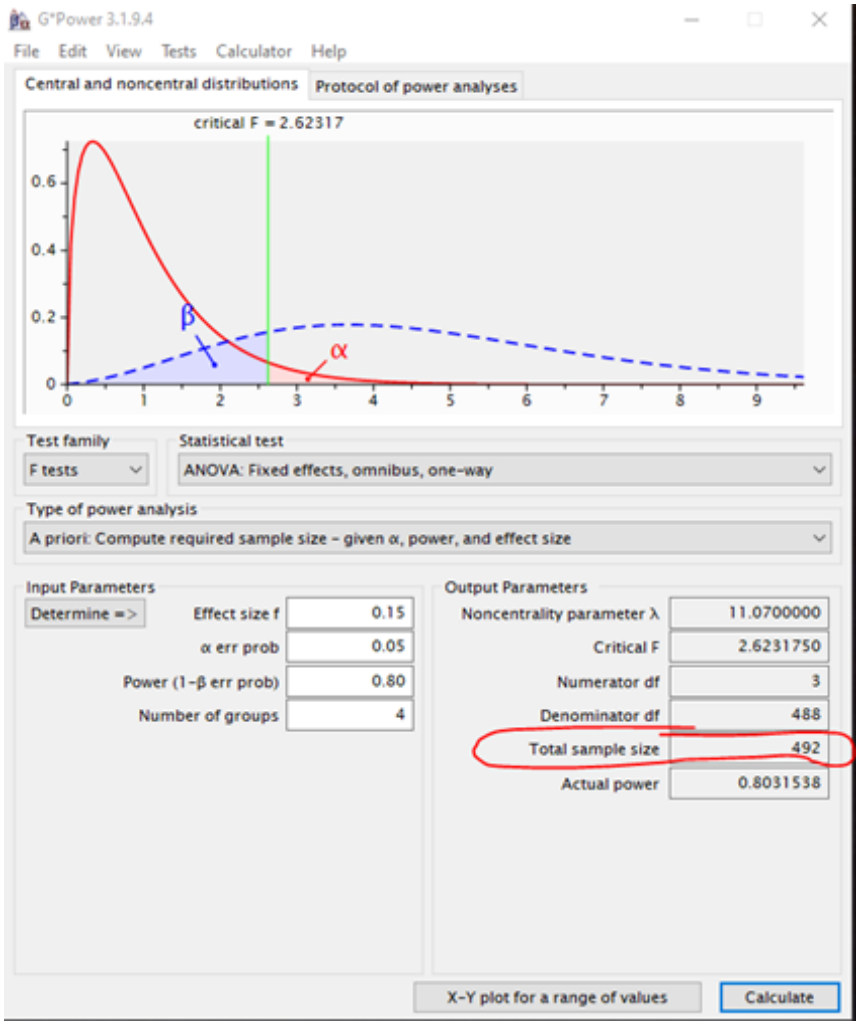
2c) Write an example of the nil hypothesis and one null hypothesis that you might have for the results of this study.

Answer:

Nil hypothesis: there is no significant difference between the mean overall GPA for students in each major.

Null hypothesis example: Psychology majors will have a mean overall GPA that is 15% higher than physics majors (many possibilities here).

3) An outside researcher heard you did a great job with your last power analysis and requested help from you. You, now a confident power analyst, agree to help. You are focused on a pilot study with 4 groups trying to see if different yoga interventions have any effect on sustained attention in a college student population. For this study, college students will be randomly assigned to each group. The plan is to conduct a one-way ANOVA. Generally, the expected effect size for interventions like these on sustained attention is 0.15. Using an alpha value of .05, what sample size would be needed to obtain a power of .80?



3a) The outside researcher mentions that they simply do not have the resources to recruit that many participants for the study. Therefore, they ask for advice on how to achieve the same power, but with less participants. What are 2 general approaches you could use to achieve this? Please explain the potential repercussions of each way described.

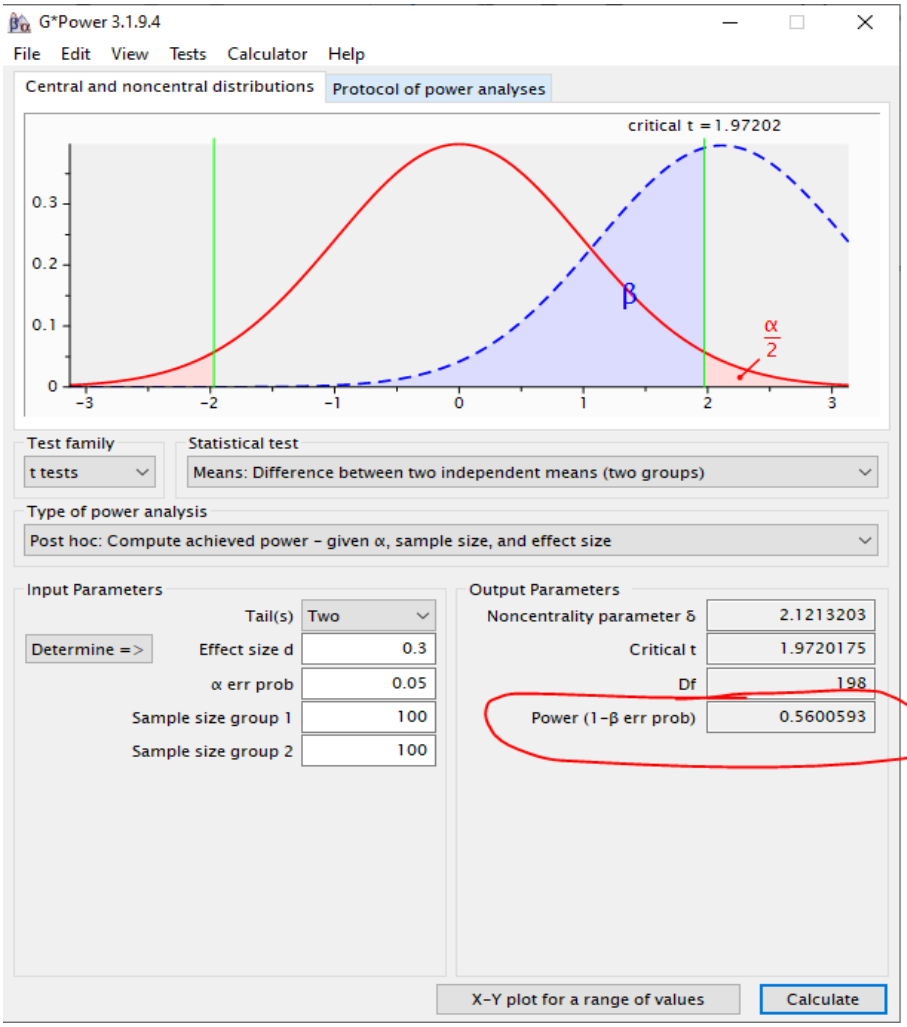
Answer:

1.) Raise the alpha level: However, this will increase the chance of a type 1 error.

2.) Make the study less complex by reducing the number of comparison conditions. If we removed one of the yoga conditions in the study (so go from 4 groups to 3 groups), the number of participants required to detect the effect will be less.

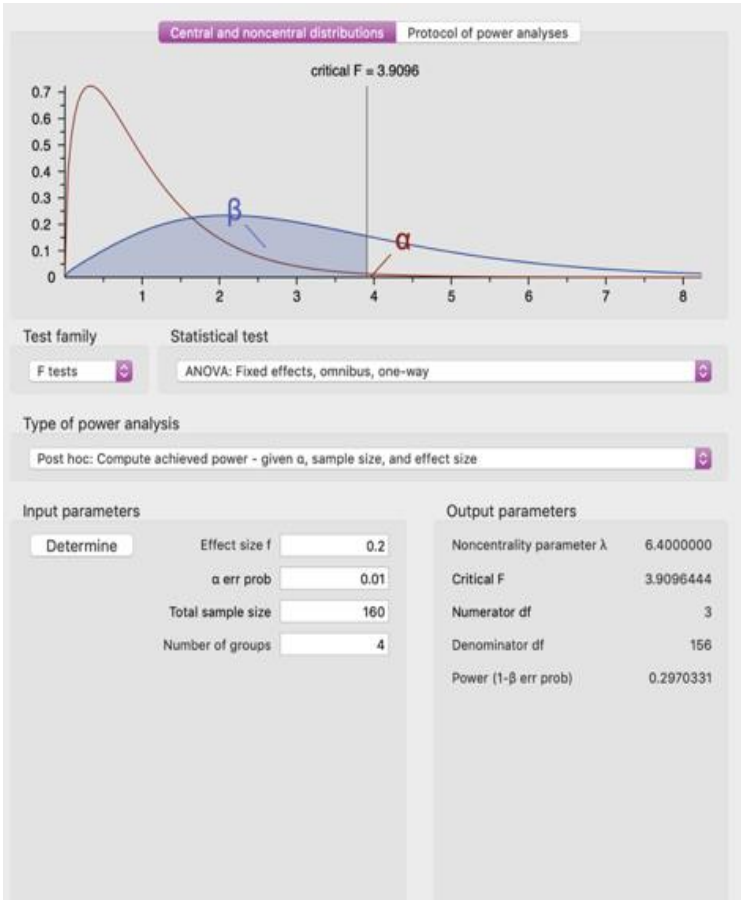
4) A CSU faculty member recently finished up an experimental study and failed to find significance between their control and treatment group when alpha was set at .05. Devastated, because they were so positive that their treatment would work, they come to you to understand why they did not find any statistical significance in their results. Their completed study yielded an effect size of .30 between the two groups. They used an independent samples t-test. The faculty members sample had 100 participants in each group. Can you provide this faculty member an explanation as to why they did not find statistically significant results, given their sample?

***"Good" Response*: Your study was underpowered, given the criteria you have provided, you only had a power level of .56, Therefore, there was only a 56% chance that you would have detected the true effect with statistical significance.**



5) You conducted a study to compare the impact of four different types of food on self-reported mood: milk chocolate, spaghetti, steamed broccoli, and a turkey sandwich. You recruited a sample of 160 undergraduate students. 40 participants were each assigned to eat one type of food (each participant only ate one of the foods) and they were asked to self-rate their mood after eating it. After you collected all of your data, you conducted a one-way ANOVA to compare the mean mood rating across the four types of food. At a significance threshold of alpha = .01, your results are non-significant. You take your results to your advisor and they ask you if your study was sufficiently powered… oops! You forgot to perform a power analysis beforehand, so you need to do it retroactively. Previous literature indicates that the effect size for the mean difference across the food groups should be around $f = 0.2$. Remembering what you learned about power in PSY 652, you would like to have a power level of at least .80. Was your study sufficiently powered to detect an effect of $f = 0.20$, if it exists, across the four groups? If not, what power level does your study currently have?

Answer: Underpowered, power = 0.297



5a) What is the probability of making a Type II error at your current power level?

Answer: Since power = 0.297, probability of making a Type II error = 0.703

5b) Fortunately, NIH thinks your research is promising and they have given you additional grant funding for this study, so you can recruit more participants. With alpha = .01, how many more people do you need to recruit to detect an effect of *f* = 0.20 if it exists in your data?

Answer: N=396 total, so need 236 more people (160 + 236 = 396)